



# Computer simulation of protein motion

W.F. van Gunsteren, P.H. Hünenberger, A.E. Mark, P.E. Smith, I.G. Tironi

Laboratory of Physical Chemistry, ETH Zürich, CH-8092 Zürich, Switzerland

Received 23 December 1994

## Abstract

The application of molecular dynamics computer simulation methods to study the dynamics of proteins is reviewed with an eye to its possibilities and limitations. Examples are given, mainly using nanosecond trajectories of the proteins bovine pancreatic trypsin inhibitor and lysozyme, of the different protein properties, of which the dynamics can be or cannot be sampled on a nanosecond time scale. It is concluded that the major asset of the simulation technique is that the different factors contributing to the dynamics of a particular process can be analyzed at atomic detail, as long as one has sampled the appropriate time scale.

*Keywords:* Protein dynamics; Molecular dynamics simulation

## 1. Introduction

Our knowledge of biomolecular systems and processes is steadily increasing due to the continuous advance of experimental techniques that reveal atomic properties of biomolecules such as proteins. X-ray diffraction provides a detailed, but static picture of the spatial atomic structure and an indication of the extent of atomic motion or disorder. Energetic information at the atomic level is largely inaccessible to experimental probes. Information concerning the dynamics of particular atoms, bond vectors or aromatic groups can be obtained by different spectroscopic techniques, including  $^{13}\text{C}$  and  $^{15}\text{N}$  nuclear magnetic resonance (NMR) relaxation, fluorescence depolarization or infrared absorption [1]. The mechanism of specific processes, such as protein folding, may be analyzed indirectly by trapping and characterizing folding intermediates [2]. Yet, the experimental characterization of the dynamics of protein atoms on the time scales ranging from femtoseconds

to seconds is far from complete, due to the difficulty of measuring dynamics into atomic detail.

An alternative method to study protein motion is to simulate protein dynamics on a computer. In the molecular dynamics (MD) simulation method, Newton's equations of motion

$$d\mathbf{r}_i(t)/dt = \mathbf{v}_i(t) \quad (1)$$

and

$$d\mathbf{v}_i(t)/dt = \mathbf{m}_i^{-1}\mathbf{F}_i(t) \quad (2)$$

with

$$\mathbf{F}_i(t) = -\partial V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)/\partial \mathbf{r}_i \quad (3)$$

for the  $N$  atoms with coordinates  $\mathbf{r} \equiv (\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ , velocities  $\mathbf{v} \equiv (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N)$  and masses  $m_i$  ( $i = 1, 2, \dots, N$ ) of a molecular system are solved numerically by integration in time  $t$ . The force  $\mathbf{F}_i$  on atom  $i$  is obtained by taking the negative gradient of the potential energy function

$$V(\mathbf{r}) = V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N), \quad (4)$$

which is assumed to represent the interaction between the atoms. The interaction function (4) is an effective interaction: it describes the interaction between the  $3N$  degrees of freedom explicitly treated in (4), averaged over the omitted atomic or electronic degrees of freedom of the real system.

Given an initial configuration  $\mathbf{r}$  of the system and initial velocities  $\mathbf{v}$  of the atoms, an MD simulation produces a series of configurations as a function of time, that is, a trajectory of the molecular system. From such a trajectory atomic or global properties of the molecular system can be calculated. If the initial configuration and velocities are representative for the equilibrium conditions (temperature, pressure, phase, external field, etc.) of the simulation, trajectory averages will represent equilibrium properties of the system. If the initial structure and velocities are not representative for the equilibrium conditions of the simulation, the structure and velocities will relax towards equilibrium. In such a non-equilibrium simulation, atomic or global properties should not be averaged, but calculated as a function of time.

Generally, MD simulation of a molecular system involves four basic choices.

(I) Which degrees of freedom of the real system are explicitly treated in (4), that is, in the simulation, and which ones are taken into account implicitly, that is, on average? For example, when simulating a protein in solution, one would like to include in the model description not only the force  $\mathbf{F}_i^{\text{int}} = -\partial V_{\text{int}}/\partial \mathbf{r}_i$  due to the explicitly simulated protein atoms, but also the external forces  $\mathbf{F}_i^{\text{ext}}$  originating from the solvent. The external force may be represented in an approximate way by its mean effect. The mean external force  $\mathbf{F}_i^{\text{mean}}$  is then e.g. the force on protein atom  $i$  due to all solvent atoms and averaged over all possible solvent configurations. The corresponding potential energy function  $V_{\text{mean}}$ , for which  $\mathbf{F}_i^{\text{mean}} = -\partial V_{\text{mean}}/\partial \mathbf{r}_i$ , is called a potential of mean force, and represents the averaged effect of the degrees of freedom which are not explicitly (implicitly) treated in the simulation. A higher-order approximation of the external force  $\mathbf{F}_i^{\text{ext}}$  is obtained by not only considering its mean effect, but also its fluctuations in time and its frictional effect:

$$\mathbf{F}_i^{\text{ext}} = -\partial V_{\text{mean}}/\partial \mathbf{r}_i + \mathbf{R}_i - m_i \gamma_i \mathbf{v}_i. \quad (5)$$

The stochastic force is denoted by  $\mathbf{R}_i$  and the frictional force is taken proportional to the instantaneous atomic velocity  $\mathbf{v}_i$  with the proportionality factor  $m_i \gamma_i$ , in which  $\gamma_i$  is the friction coefficient, which depends on the solvent viscosity and on the degree of contact between protein atom and solvent. At least those degrees of freedom that are essential for a proper representation of the dynamical quantities one wishes to simulate and study, should be explicitly modelled.

(II) Which interaction function or force field  $V(\mathbf{r})$  is used to calculate the potential energy of the system and the forces along the explicitly treated degrees of freedom? The typical energy function (4) for a biomolecular system contains a variety of terms representing covalent bonding, bond-angle, torsional angle, van der Waals and Coulombic interactions. Many-body terms, e.g. to model atomic polarization, or time-dependent terms representing experimental information, such as scattering or absorption intensities, on the specific molecular system that is simulated, may be included in (4) too [3,4]. In summary,  $V(\mathbf{r})$  must represent the energy landscape of the molecular system that is to be simulated. The shape of this landscape in regions that are accessible at the thermodynamic state point of interest, will determine the dynamical behaviour of the simulated system.

(III) Which equations of motion are integrated to simulate the behaviour of the molecular system? Eqs. (1) and (2) are Newton's equations of motion, valid for Cartesian coordinates and classical mechanics. Alternative formulations of classical mechanics, such as the equations of motion of Lagrange or of Hamilton, may be used too [3]. If the explicitly treated degrees of freedom are essentially of quantum-mechanical nature, the Schrödinger equation should be integrated [5]. If the atomic forces  $\mathbf{F}_i$  depend on the atomic velocities, as in (5), Eq. (2) is transformed into the stochastic Langevin equation

$$d\mathbf{v}_i(t)/dt = m_i^{-1} [\mathbf{F}_i^{\text{int}}(t) + \mathbf{F}_i^{\text{mean}}(t) + \mathbf{R}_i(t)] - \gamma_i \mathbf{v}_i(t). \quad (6)$$

If the inertial term on the left hand side of (6) is small compared to the force terms on the right hand side, equation of motion (6) simplifies to

$$\mathbf{v}_i(t) = (m_i \gamma_i)^{-1} [\mathbf{F}_i^{\text{int}}(t) + \mathbf{F}_i^{\text{mean}}(t) + \mathbf{R}_i(t)], \quad (7)$$

which forms together with Eq. (1) the equations for Brownian motion [3]. It goes without saying that only dynamical simulation techniques, that is, methods which integrate equations of motion, can be used to study dynamical properties of molecular systems, in contrast to Monte Carlo methods.

(IV) Which spatial or thermodynamic boundary conditions are used in the simulation? When simulating a system of finite size ( $N \ll$  Avogadro's number), some thought must be given to the way the spatial boundary of the system will be treated. The simplest choice is the vacuum boundary condition, which mimics a gas-phase environment. When this boundary condition is used for a protein in solution, the dynamics of atoms near or at the surface of the system will be distorted, and the dynamics of global system properties will not represent those of the condensed phase. Solvating the protein in a sphere of explicit solvent molecules only shifts the distortive boundary effects from the solute–solvent to the solvent–vacuum interface. The structural deformations due to the vacuum outside the molecular system may be reduced by designating a layer of atoms of the system to a so-called extended wall region, in which the motion of the atoms is restrained in order to reduce the deforming influence of the nearby vacuum. The atoms in the extended wall region can be kept fixed or harmonically restrained to stationary positions. Although the extended wall boundary condition may preserve the structural properties of the system, the restraining will generally strongly influence the dynamics of the system, especially of the more global modes. To minimize the distortive effects of the spatial boundaries on the dynamics of a molecular system in the condensed phase, the best method is the use of periodic boundary conditions, in which the simulated system is surrounded by identical translated images of itself [3]. The application of other types of restraints or constraints, such as bond-length or bond-angle constraints, atom–atom distance restraints derived from NMR spectroscopic data or structure factor restraints derived from X-ray diffraction data, will also distort the dynamical properties of particular physical quantities [4,6]. Thermodynamic restraints or constraints, such as the instantaneous or mean coupling of the system to a temperature bath or pressure bath, will influence the fluctu-

ations in the conjugate quantities, such as the energy or volume of the system [7,8]. When studying the dynamics of a molecular system by simulation, one should be aware of the distortive effect the use of constraints or restraints of spatial or thermodynamic nature will have on the dynamical behaviour of atomic and system properties.

The central question when using MD simulation to study the motion of molecular systems such as proteins is: how close does a computer-generated trajectory come to that of a real protein? The quality of an MD simulation is limited by the following five factors.

(A) Can the degrees of freedom that are not taken into account in the molecular model reliably be neglected, or included on average in a potential of mean force?

(B) Is the size of the simulated system sufficiently large to avoid distortion of the dynamics by the unphysical spatial boundary conditions? Will the thermodynamic or structural constraints or restraints distort the dynamics and fluctuations of particular atomic or system properties?

(C) Are the type, functional form and parameters of the effective interaction function  $V(\mathbf{r})$  of sufficient quality to reproduce the dynamical behaviour sufficiently accurately?

(D) If the motion is classically simulated, can quantum-mechanical effects reliably be neglected?

(E) Is the length of the simulation sufficiently long to yield reliable trajectory averages of the different molecular or system properties? Or, is the simulation length considerably longer than the relaxation time of the property of interest?

The answers to these five basic questions will vary with the type of molecular system and the type of property of this system one is interested in. Here, we will concentrate on the fifth question with respect to a variety of properties of proteins in solution.

The time scale of the dynamics of different properties of proteins ranges from femtoseconds to seconds or even longer. Due to limited computing power, MD simulation of a relatively small protein in aqueous solution at the classical, atomic level currently reaches the nanosecond time scale. For simplified protein models in which amino acid residues are

taken as particles and the influence of solvent is approximated by a mean force, the microsecond time scale can be reached. This implies that many properties of interest that have a relaxation time longer than the simulation time can in general not be studied by simulation. It is the purpose of this paper to review which protein motions can be and which cannot be reliably sampled in an MD simulation.

The analysis of the time scale of different types of motion of a protein in a simulation may serve two purposes.

(1) One is interested in the protein dynamics per se, e.g. how it depends on the shape of the potential energy landscape (4) or which degrees of freedom are important to a particular process.

(2) In order to reliably interpret a simulation one should estimate the accuracy of the different simulated properties. Trajectory averages will generally only be representative when the equilibration time of the simulation,  $\tau_{\text{equil}}$ , is longer than the relaxation time  $\tau_{\text{relax}}(Q)$  of the property  $Q$ ,

$$\tau_{\text{equil}} > \tau_{\text{relax}}(Q), \quad (8)$$

and when the sampling time,  $\tau_{\text{sample}}$ , is much longer than  $\tau_{\text{relax}}(Q)$ ,

$$\tau_{\text{sample}} \gg \tau_{\text{relax}}(Q). \quad (9)$$

If conditions (8) or (9) are not fulfilled, the trajectory average  $\langle Q(t) \rangle$  of the property  $Q$  will display a drift as a function of time or erratic behaviour due to the occurrence of rare events affecting the value of  $Q(t)$ . Below, we shall give examples of such cases.

The relaxation time  $\tau_{\text{relax}}(Q)$  may be long for different reasons:

(a) The system may change relatively rarely, but in a fast manner, between relatively stable states. An example is the flipping of Phe sidechains in a protein which is a fast, picosecond time-scale process, which occurs comparatively infrequently, only on a millisecond time scale. In such a case the trajectory averages will be sensitive to the number of rare events that are simulated.

(b) The system may change intrinsically slow, in which case trajectory averages will display a continuous change as a function of time.

The relaxation and dynamics of the different protein properties occurring in an MD simulation can be analyzed by different means.

(1) For equilibrium simulations one may monitor the time series of a property  $Q(t)$ , or its average  $\langle Q(t) \rangle$  or fluctuations  $\langle [Q(t) - \langle Q \rangle]^2 \rangle^{1/2}$ , or calculate its autocorrelation function  $\langle Q(t')Q(t'+t) \rangle$ . The decay time of the auto-correlation function, or the build-up rates of the trajectory averages give an indication of  $\tau_{\text{relax}}(Q)$ .

(2) When starting a simulation from a non-equilibrium initial state, the rate of relaxation towards equilibrium for different properties  $Q(t)$  will give an indication of  $\tau_{\text{relax}}(Q)$ .

(3) If different MD simulations starting from different initial states do not converge to the same trajectory average for property  $Q$ , it can be concluded that  $\tau_{\text{relax}}(Q)$  is longer than the simulation time.

In this paper we give examples of the way protein motion is reflected in the various protein properties that can be calculated from simulated trajectories. In Section 2, we briefly characterize the different types of motion and their dependence on thermodynamic parameters. Subsequently, we demonstrate the dependence of different energetic and structural quantities upon the time scale of protein and solvent motion. Out of convenience the examples have been taken from our own current and previous work. In Section 3 we discuss the results and draw a few practical conclusions with respect to the simulation and analysis of protein motion.

## 2. Time scales of molecular motion

The time scale of change or relaxation of a particular physical quantity calculated for a particular molecular system will depend on the type of molecular system, the thermodynamic state point of interest and the particular quantity or property.

### 2.1. Type of molecular system

The dynamics of a protein in solution will involve different types of motions. For the protein we may distinguish overall translation and rotation, relative

motion of domains, secondary structure elements or loops, the motion of amino acid residue sidechains, flips of backbone ( $\varphi$ ,  $\psi$ ,  $\omega$ ) or sidechain ( $\chi$ ) angles, internal vibrations of secondary structure elements, and bond-angle and bond-length vibrations. The characteristic times of these motions range from femtoseconds to tens of nanoseconds, and can therefore only partially be sampled in a nanosecond trajectory. For the solvent the relaxation times cover a smaller range, from femtoseconds to tens of picoseconds for collective properties, which can be properly sampled when averaging over many (hundreds) of solvent molecules and nanosecond trajectories.

## 2.2. Thermodynamic state point

The extent and relaxation time of the motion of a molecular system will be very sensitive to the thermodynamic state point at which the simulation is carried out. For protein crystals below about 250 K the atomic motion is mainly harmonic. At room temperature a folded protein will display very non-harmonic motion around its average folded conformation, which also will depend on its solvent environment [9]. At high pressure, the motions seem slightly slowed down [10]. The effects of a change in pH or ionic strength of the solution upon the dynamics of a protein have not yet been studied in detail by MD simulation.

## 2.3. Property of interest

In this section we illustrate the differences in relaxation of different physical quantities as observed in MD simulations of the proteins hen egg-white lysozyme and bovine pancreatic trypsin inhibitor (BPTI), and of the cyclic sugar molecule  $\alpha$ -cyclodextrin in aqueous solution. The relaxation behaviour of a pure liquid is illustrated using chloroform as example and that of a high salt solution by considering a solution of 1 molar NaCl. The set-up of each of these simulations has been reported elsewhere [10–18] and will not be repeated here. All simulations were carried out using the GROMOS simulation package and force field [19], in some cases with the modification mentioned in [17].

### 2.3.1. Energy

When simulating a system in equilibrium various properties such as the total potential energy or the different types of energies (Coulomb, van der Waals, etc.) are generally monitored in order to determine when the system reaches a more or less stable state. However, a stable behaviour of these types of total potential energies does not necessarily imply that the energy components due to different parts of the system are at equilibrium. This is illustrated in Fig. 1 for an MD simulation of lysozyme in water. The total van der Waals energy seems to be stabilized after about 500 ps (upper panel), but the protein–water van der Waals energy is steadily decreasing accompanied by a steady (smaller) increase of the protein–protein van der Waals energy (lower panel) and the water–water energy (not shown), during the whole 1 ns simulation period. The slow relaxation of the protein–water interface is only apparent if the appropriate energy component is monitored.

### 2.3.2. Structural quantities

A variety of structural properties may be monitored as a function of time. Fig. 2 shows the root mean square (rms) deviation of the  $C_\alpha$  atom positions averaged over 50 ps for three MD simulations of lysozyme from their initial positions in the X-ray

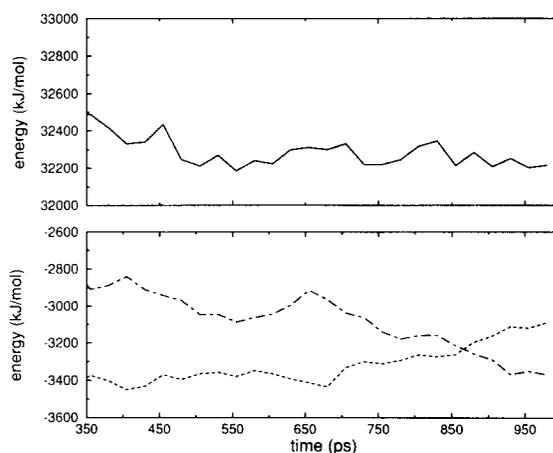


Fig. 1. Van der Waals non-bonded energy of lysozyme in water as a function of time. The data points represent 25 ps averages. Upper panel: total energy (solid line). Lower panel: intra-protein energy (dotted line), protein–water energy (dot–dashed line).

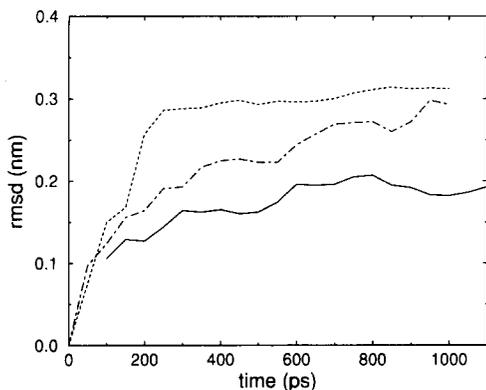


Fig. 2. Root mean square positional deviation of the  $C_{\alpha}$  atoms between the initial X-ray structure and 50 ps averages from three MD simulations of lysozyme. Dotted line: lysozyme in vacuo, GROMOS87, 37D4 force field [19]. Dot-dashed line: lysozyme with 5345 SPC water molecules, GROMOS87, 37C4 force field [19]. Solid line: idem, but with modified carbon-to-water-oxygen van der Waals parameter [17].

structure. The deviation from the X-ray structure stabilizes much faster in vacuo (dotted line), after about 250 ps, than in aqueous solution. Using the standard GROMOS87 force field with the simple point charge (SPC) water model [19] the deviation is steadily increasing over a whole nanosecond (dot-dashed line). This is due to a too favourable carbon-to-water-oxygen van der Waals interaction, which tends to slowly expose more and more hydrophobic area to the solvent. When this interaction is modified [17], the deviation from the X-ray structure stabilizes after about 600 ps (solid line). Fig. 2 illustrates that the structural relaxation of a protein is much slower in water than in vacuo.

Conventional crystallographic refinement of a protein structure also yields a set of isotropic atomic temperature or B-factors. Their size is indicative of the amount of motion or disorder that is present in the crystal on the time scale of the diffraction experiment, which ranges from seconds to days. Using the relation

$$B_i = (8\pi^2/3)\Delta r_i^2, \quad (10)$$

where  $\Delta r_i^2$  is the mean square fluctuation of atom  $i$  around its average position, one may calculate the correlation between the crystallographic atomic B-factors and the ones obtained by averaging over different periods of an MD trajectory. The correla-

tion coefficients obtained for the MD simulation of lysozyme in water using the modified GROMOS force field [17] are shown in Fig. 3. The curves based on B-factors averaged over 50 ps or 200 ps show that these averaging periods are too short to obtain a reliable estimate of the atomic positional fluctuations. Only when averaging over 500 ps or longer do the B-factors become less sensitive to the time points between which the averaging is carried out. Fig. 3 illustrates that the build-up time of atomic positional fluctuations in proteins is of the order of hundreds of picoseconds.

Hydrogen bonds are formed and broken on all time scales observable in an MD simulation. In crystals of  $\alpha$ -cyclodextrin-6H<sub>2</sub>O, two water molecules, denoted by A and B, are enclosed in the  $\alpha$ -cyclodextrin cavity. Neutron diffraction results indicate that the predominant hydrogen bonding configuration of these water molecules is OWA-H...OWB [21]. A 15 ps MD simulation of 4 unit cells of crystalline  $\alpha$ -cyclodextrin, comprising 16  $\alpha$ -cyclodextrin molecules, has been reported earlier [11]. Fig. 4 shows the direction of the hydrogen bond between the 16 equivalent pairs of water molecules A and B, and which H-atom is involved in the hydrogen bond, as a function of time. Only 6 pairs

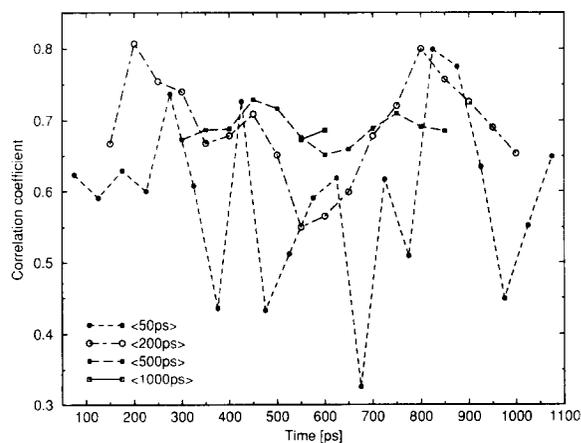


Fig. 3. Correlation coefficient for atomic B-factors or mean square fluctuations (Eq. (10)) obtained by conventional crystallographic refinement of lysozyme [20] and calculated over different periods of the MD simulation of lysozyme in water with the modified GROMOS force field [17]. The averaging windows are: 50 ps (dotted line), 200 ps (dot-dashed line), 500 ps (dashed line) and 1000 ps (solid line).

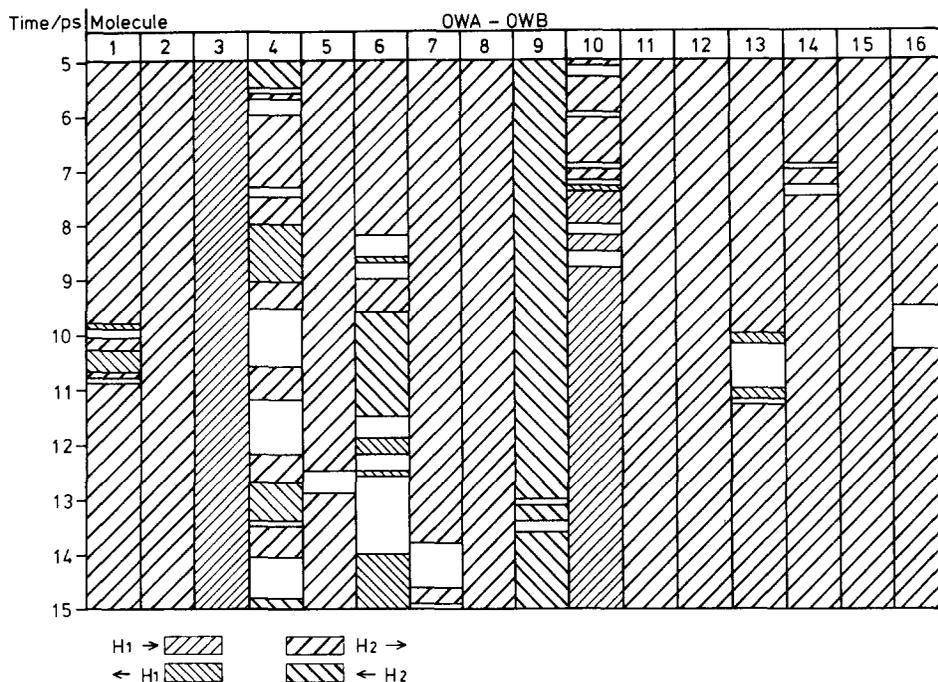


Fig. 4. Presence and directionality of the hydrogen bonds between water molecules A and B in the cavities of the 16  $\alpha$ -cyclodextrin molecules of 4 unit cells of crystalline  $\alpha$ -cyclodextrin-6H<sub>2</sub>O during a 15 ps MD simulation [11].

show uninterrupted hydrogen bonds. Five pairs show flip-flop behaviour, i.e. the hydrogen bond changes direction. On average the OWA-H...OWB configuration is present for 80% of the time, in accordance with experiment [11]. The type of behaviour shown in Fig. 4 is also observed in much longer MD simulations of proteins.

Fig. 5 illustrates how rare events can influence a trajectory average calculated from an MD simulation. The upper panel shows the value of the backbone  $\varphi$ -angle (C-N-C <sub>$\alpha$</sub> -C) of residue Ile 78 of lysozyme as a function of time for the MD simulation of lysozyme in water using the modified GRO-MOS force field. The  $\varphi$ -angle switches occasionally between two relatively stable states. The order parameter

$$S^2 = \frac{1}{2} \left( 3 \sum_{\alpha=1}^3 \sum_{\beta=1}^3 \langle \mu_{\alpha} \mu_{\beta} \rangle^2 - 1 \right), \quad (11)$$

which can be related to <sup>15</sup>N NMR relaxation parameters, is sensitive to the value of the  $\varphi$ -angle, since  $\mu_{\alpha}$  ( $\alpha = x, y, z$ ) denotes the three cartesian components of the <sup>15</sup>N-H vector of the peptide nitrogen. The

lower panel of Fig. 5 shows the trajectory averaged  $S^2$  parameters as a function of time for three residues, Cys 76 (dotted line), Asn 77 (dashed line) and Ile 78 (solid line).

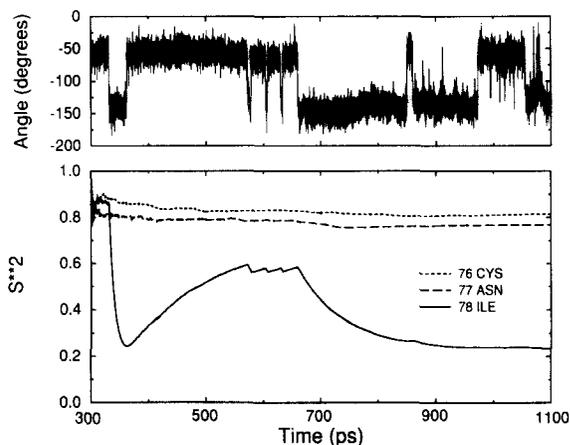


Fig. 5. Peptide  $\varphi$ -angle of Ile 78 of lysozyme as a function of time for the period 300–1100 ps of the MD simulation of lysozyme in water using the modified GROMOS force field. The lower panel displays the <sup>15</sup>N order parameters  $S^2$  (see Eq. (11)) for residues Cys 76 (dotted line), Asn 77 (dashed line) and Ile 78 (solid line).

(solid line). If the averaging period is of the same order of magnitude as the time between the rare  $\varphi$ -angle flips, every  $\varphi$ -flip is reflected in a change in  $S^2$ . So, for a proper evaluation of the accuracy of trajectory averages, the analysis of averages as a function of time is a necessary, but not sufficient condition.

The conformation of a protein is generally characterized in terms of secondary structure elements, such as  $\alpha$ -helices,  $\beta$ -sheets, etc. In Fig. 6 the evolution of different secondary structure elements of lysozyme in water is displayed as a function of time. The  $\beta$ -sheet and the  $\alpha$ -helices A, B and C are rather stable, although some fraying is observed at the ends of the helices. The  $3^{10}$  helices change character to  $\alpha$ -helical, only very rarely recurring to a  $3^{10}$  form. Fig. 6 illustrates the fluctuating nature of protein structure, but does not indicate the occurrence of more collective modes.

Since the native state of lysozyme is characterized by two structural domains, one consisting primarily of  $\alpha$ -helices ( $\alpha$ -domain, residues 1–37 and 88–129) and the other dominated by a section of triple-

stranded  $\beta$ -sheet ( $\beta$ -domain, residues 41–84), one might expect to find a low-frequency vibration of the two domains with respect to each other. The autocorrelation function for the distance between the centres of mass of the two domains as calculated from the trajectory of lysozyme in water with the non-modified GROMOS force field is shown in Fig. 7. It displays a damped oscillatory motion with a frequency of about  $0.08 \text{ cm}^{-1}$  and a friction coefficient of  $2.3 \text{ ns}^{-1}$ . However, since the length of the MD simulation is only slightly more than twice the oscillation time of 400 ps, the statistical reliability of the slow mode in Fig. 7 is virtually zero. The part of the correlation function beyond 100 ps, a tenth of the simulation period, may not be considered to be statistically well sampled. So, no conclusions about hinge-bending motion in lysozyme should be drawn from Fig. 7.

Other slow processes are the translational and rotational motion of a protein in solution. These motions have been analyzed using 1 ns trajectories of BPTI and lysozyme in water [13]. The translational diffusion constants were in good agreement

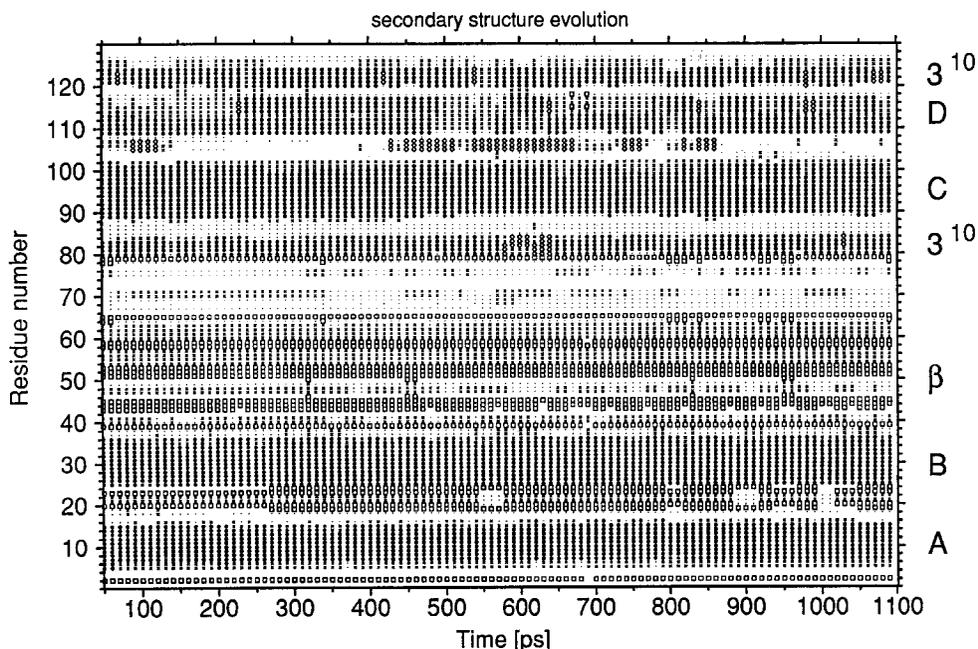


Fig. 6. Secondary structure of lysozyme in water as a function of time, as defined by the "SUMMARY" entry of the DSSP program [22]. ●:  $\alpha$ -helix, □:  $\beta$ -bridge or  $\beta$ -sheet, ◇:  $3^{10}$ -helix, ◆:  $I$ -helix, x: hydrogen-bonded turn and ·: bend.

with existing experimental values, but the rotational correlation times were slightly short, indicating faster rotational diffusion compared to experiment.

### 2.3.3. NMR relaxation

NMR experiments can provide valuable information concerning the dynamics of proteins in solution. The dipolar relaxation between two nuclei can be related to the time correlation function  $C(t)$  describing the orientation of the interatomic vector between the two nuclei [23]. If the relaxation due to the overall rotation of the protein is decoupled from the relaxation due to internal dynamics, the total correlation function can be factorized,

$$C(t) = \frac{1}{5}C_R(t)C_1(t), \quad (12)$$

where the correlation functions  $C_R(t)$  and  $C_1(t)$  are related to the overall rotational tumbling and the internal dynamics of the protein. Assuming isotropic Debye rotational diffusion of the protein we have

$$C_R(t) = e^{-t/\tau_R}, \quad (13)$$

where  $\tau_R$  is the rotational correlation time.

The internal correlation function is given by [24]

$$C_1(t) = \langle P_2(\mu(0) \cdot \mu(t)) \rangle, \quad (14)$$

where  $\mu(t)$  is the orientation of the interatomic vector at time  $t$  as measured in the protein coordi-

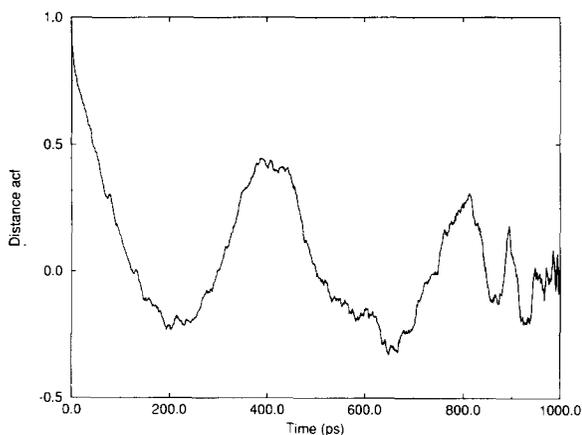


Fig. 7. Autocorrelation function of the distance between the centres of mass of the  $\alpha$ -domain (residues 1–37 and 88–129) and of the  $\beta$ -domain (residues 41–84) of lysozyme for a 1000 ps MD simulation of this protein in water, using the non-modified GRO-MOS force field.

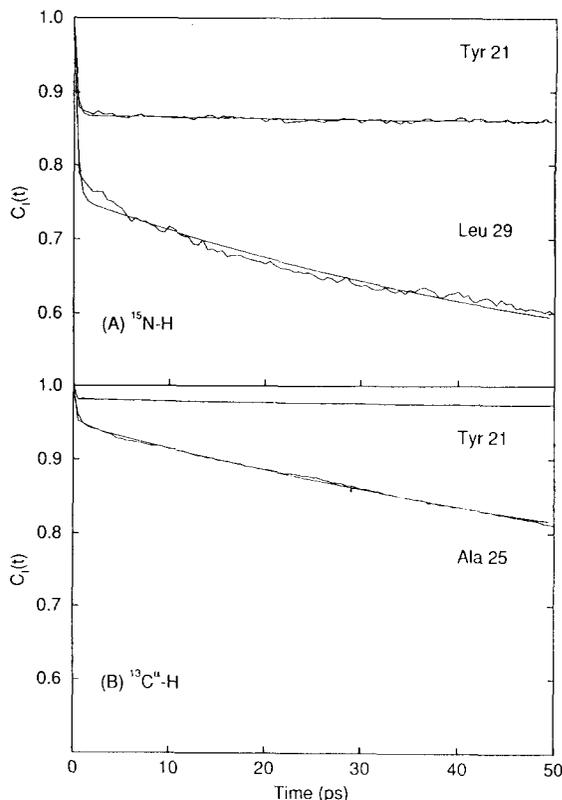


Fig. 8. Internal orientational correlation functions  $C_1(t)$  (Eq. (14)) for the  $^{15}\text{N-H}$  (A) and  $^{13}\text{C}_\alpha\text{-H}$  (B) vectors obtained from a 1 ns MD trajectory of BPTI in a box with 2371 water molecules [15]. A bi-exponential fit to the functions is also shown.

nate frame,  $P_2$  is the 2nd rank Legendre polynomial and  $\langle \dots \rangle$  denotes a trajectory average over the simulation. Typical examples of the correlation function  $C_1(t)$  obtained for the N-H and  $\text{C}_\alpha\text{-H}$  vectors of different residues of BPTI from a 1 ns MD simulation of this protein in a box with 2371 water molecules, are shown in Fig. 8 [15]. The correlation functions display a fast initial decay with only a small additional decay from longer time motions. In general, the N-H vectors display a higher degree of motional averaging than the  $\text{C}_\alpha\text{-H}$  vectors.

Fig. 9 shows the  $T_1$  relaxation rates for BPTI as a function of residue number as obtained from the MD simulation [15]. The solid line represents the contribution from the overall rotation of the protein, the dotted line represents the contribution from the slow internal decay processes. The contribution of the fast

internal decay processes is negligibly small. The calculated  $T_1$  relaxation rates are determined almost solely by the order parameters ( $S^2$ ) and the overall rotational correlation time ( $\tau_R$ ) of the protein, which quantities are poorly sampled even in a 1 ns MD simulation.

### 2.3.4. Dielectric relaxation

The dielectric properties of a protein can be related [25] to the fluctuations and the autocorrelation function of the total dipole moment of the protein

$$\mathbf{M}(t) = \sum_{i=1}^N q_i \mathbf{r}_i(t), \quad (15)$$

where  $q_i$  is the charge on atom  $i$ ,  $\mathbf{r}_i$  is the position of atom  $i$  with respect to the centre of mass of the protein, and  $N$  is the number of protein atoms. The

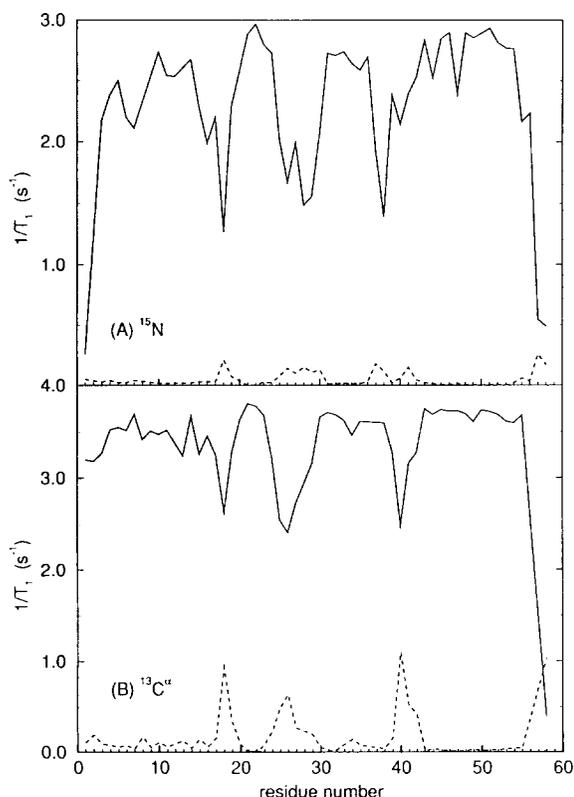


Fig. 9. Contributions to the  $T_1$  relaxation rates for the  $^{15}\text{N}$  (A) and  $^{13}\text{C}_\alpha$  (B) nuclei of BPTI obtained from a 1 ns MD trajectory in water [15] as a function of residue number. Solid line: contribution from the overall rotation of the protein. Dotted line: contribution from the slow internal decay processes. The contribution from the fast internal decay is negligibly small.

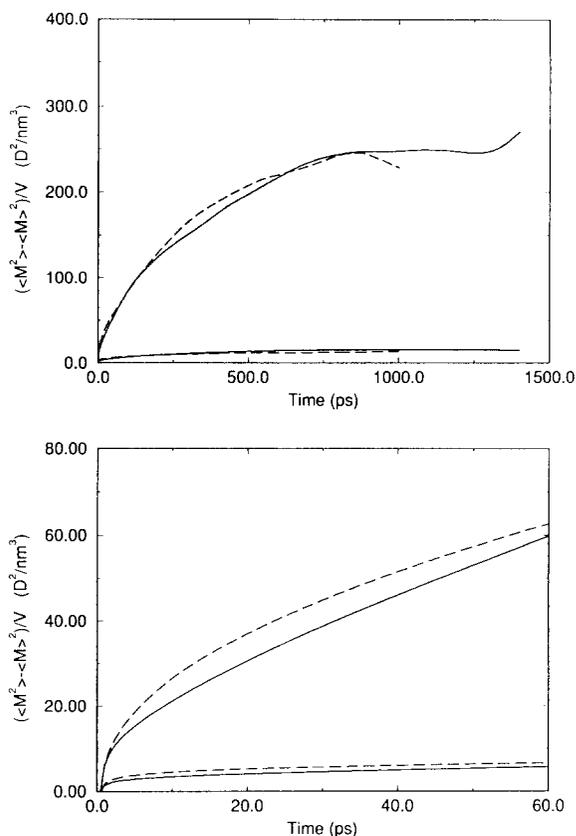


Fig. 10. Protein dipole moment (Eq. (15)) fluctuation densities obtained from MD simulations of BPTI (solid line) and lysozyme (dashed line) in water, using the non-modified GROMOS force field [12]. The upper set of curves represents fluctuations for all atoms while the lower curves are the fluctuations of just the peptide (CO–NH) groups.

dipole moment fluctuation densities for the proteins BPTI (solid line) and lysozyme (dashed line) as obtained from long MD simulations of these proteins in water are shown in Fig. 10 [12]. The upper set of curves represents fluctuations for all atoms while the lower curves are the fluctuations of just the peptide (CO–NH) groups. The peptide groups display a significantly smaller equilibrium fluctuation density due to their restricted motion as compared with sidechain motion. For both proteins the fluctuation density converges in about 1 ns. This is a much longer build-up time than is generally observed for a liquid. Fig. 11 shows the total dipole moment fluctuations as a function of time calculated from an MD trajec-

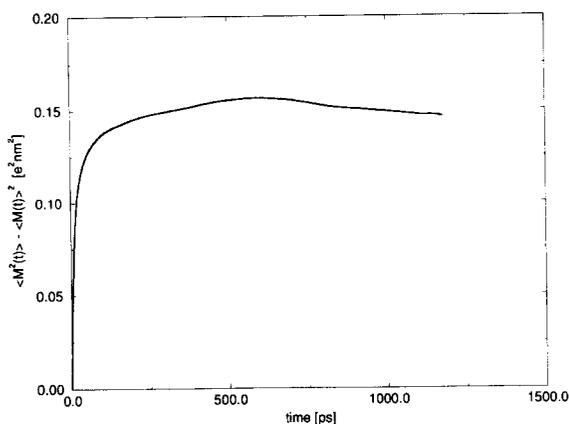


Fig. 11. Total dipole moment fluctuation of a box with 216 chloroform molecules obtained from MD simulation [16].

tory for a box with 216 chloroform molecules [16]. It displays a faster build-up.

The autocorrelation function

$$\Phi(t) = \langle \mathbf{M}(0) \cdot \mathbf{M}(t) \rangle / \langle M^2(0) \rangle \quad (16)$$

of the total dipole moment  $\mathbf{M}$  gives insight into the dynamics of the dielectric medium. Fig. 12 shows this function for BPTI (solid line) and lysozyme (dashed line) [12], while Fig. 13 contains  $\Phi(t)$  for liquid chloroform [16]. The dielectric relaxation time of the latter is of the order of picoseconds, whereas for the proteins BPTI and lysozyme it is of the order of nanoseconds. This reflects the different time scales governing the atomic motions in these molecular systems.

### 2.3.5. Free energy of complex formation

The free energy of a molecular system is dependent on the extent of phase or configuration space that is accessible to the system at the thermodynamic state point of interest. Complete sampling of configuration space is not possible except for the simplest of model systems. Free-energy calculations of biomolecular systems have thus concentrated on the determination of the relative free energy between two closely related states. In this way only differences between the two states need be considered. Irrelevant regions of configuration space can be ignored. Although this greatly simplifies the problem, it does not eliminate the necessity to sample relevant regions of configuration space, especially if entropic

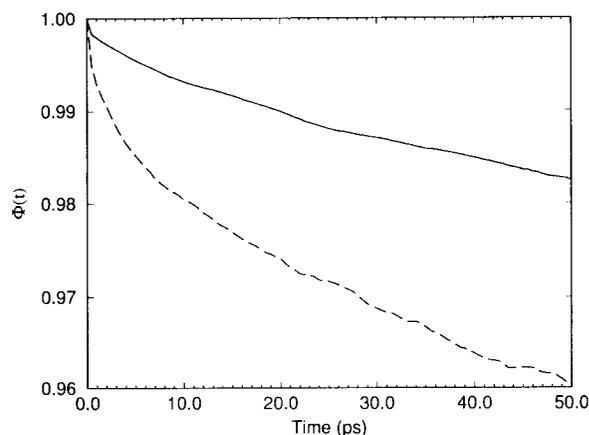
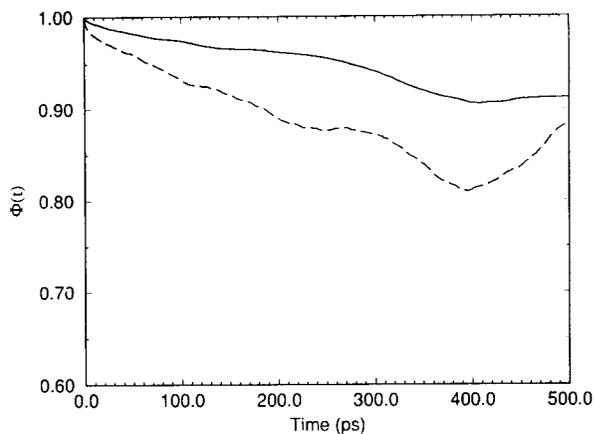


Fig. 12. Protein dipole moment autocorrelation functions (Eq. (16)) obtained from MD simulations of BPTI (solid line) and lysozyme (dashed line) in water [12].

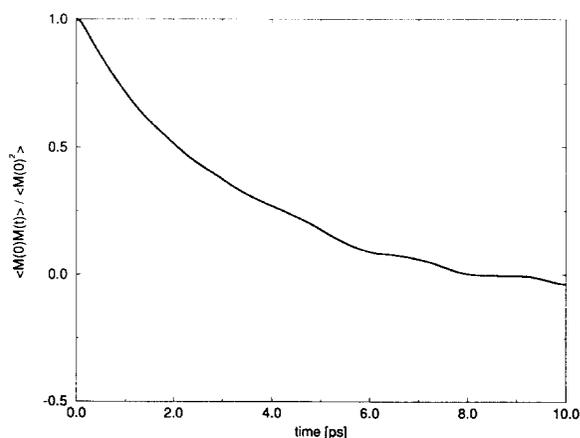


Fig. 13. Total dipole moment autocorrelation function of a box with 216 chloroform molecules obtained from MD simulation [16].

contributions to the free energy are to be correctly estimated. If a molecular system shows motions on different time scales, these may contribute differently to the free energy.

The sensitivity of the free energy calculated from an MD trajectory to the length of the trajectory has been demonstrated for the case of the relative free energy of binding of p-chlorophenol versus p-methylphenol in the cavity of  $\alpha$ -cyclodextrin in aqueous solution [17]. The difference in free energy between two states A and B of a system, of which the interaction functions (4) are denoted by  $V_A$  and  $V_B$ , can be calculated from the expression

$$\Delta G_{BA} = \int_{\lambda_A}^{\lambda_B} \langle \partial V / \partial \lambda \rangle_{\lambda} d\lambda. \quad (17)$$

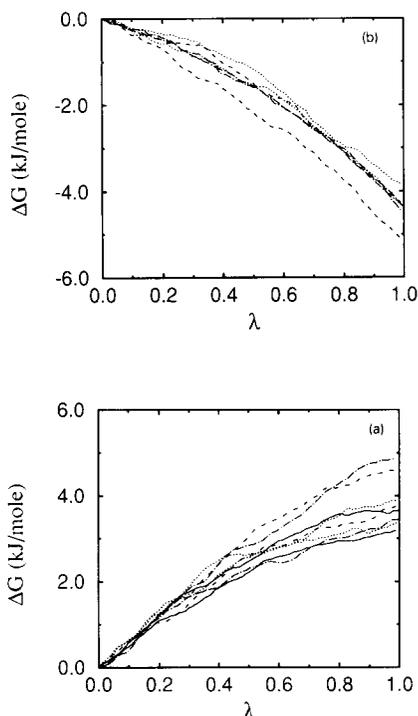


Fig. 14. Change in free energy (Eq. (17)) as a function of the coupling parameter  $\lambda$  for the mutation of p-chlorophenol ( $\lambda_A = 0$ ) to p-methylphenol ( $\lambda_B = 1$ ) in water (A) and when bound inside the cavity of  $\alpha$ -cyclodextrin (B) [17]. The different line styles represent MD simulations of different lengths: 25 ps (dotted line), 50 ps (dashed line), 100 ps (dot-dashed line), 300 ps (solid line). The curves have been shifted vertically such that  $\Delta G = 0$  at  $\lambda = 0$  for ease of comparison.

Table 1

Calculated free energies for the mutation from p-Cl-phenol to p-CH<sub>3</sub>-phenol. Values are in kJ mole<sup>-1</sup>

time (ps)	water		$\alpha$ -CD		$\alpha$ -CD-water
	$\Delta G$	hysteresis	$\Delta G$	hysteresis	
25	-4.33	0.25	3.68	0.51	8.0
50	-4.90	1.22	4.16	0.77	9.1
100	-4.50	0.21	4.18	1.48	8.7
300			3.42	0.51	7.9 <sup>a</sup>

<sup>a</sup> Value calculated using the  $\Delta G$  for water calculated over 100 ps.

The potential energy function  $V$  is made a function of the coupling parameter  $\lambda$ , such that  $V(\lambda_A) = V_A$  and  $V(\lambda_B) = V_B$ . Averaging over configurations generated with the interaction function  $V(\lambda)$  is denoted by  $\langle \dots \rangle_{\lambda}$ . The coupling parameter can be made a function of time,  $\lambda(t)$ , such that it slowly changes from  $\lambda_A$  to  $\lambda_B$  over the time course of an MD simulation.

Fig. 14 shows the change in free energy as a function of the coupling parameter  $\lambda$  for the mutation of p-chlorophenol ( $\lambda_A = 0$ ) to p-methylphenol ( $\lambda_B = 1$ ) in water (A) and when bound in the cavity of  $\alpha$ -cyclodextrin in water (B). The different line styles represent simulations of different lengths. Each change has been carried out both in forward and reverse directions. The difference between  $\Delta G_{BA}$  and  $-\Delta G_{AB}$  is called the hysteresis, which should be zero if the change is carried out in a reversible manner. The average changes in free energy for the forward and reverse mutations and the hysteresis are given in Table 1. The hysteresis is a non-monotonic function of time, which is due to different relaxation processes contributing to  $\Delta G$ . When the simulation time is of the same size as the relaxation time of a particular mode of motion, the hysteresis will be larger than when these times are very different. So, monitoring of the hysteresis as a function of time in a free energy simulation gives an indication about the time scales of the different modes that are successively accessed when the simulation is proceeding. Fig. 15 illustrates the range of motion of the guest molecule within the  $\alpha$ -cyclodextrin molecule.

### 2.3.6. Solvent dynamics

When simulating a protein in solution the dynamics of the solvent molecules near the protein surface

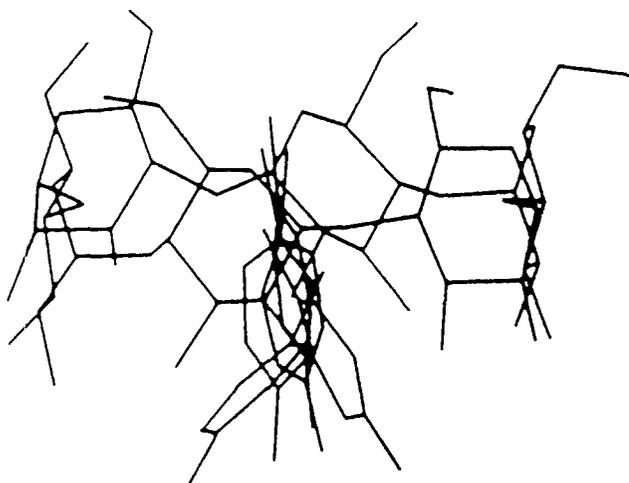


Fig. 15. Superposition of five snapshots from an MD trajectory of *p*-chlorophenol bound to  $\alpha$ -cyclodextrin showing the range of motion of the guest within the host molecule [17].

may be analyzed and correlated with the type of protein surface. Such an analysis has been carried out for a 1 ns MD simulation of BPTI in aqueous solution [14]. Fig. 16 shows a few typical residence time distributions for the time a water molecule resides in the first hydration shell of a particular protein atom. The residence times of individual water molecules coming near a given BPTI atom vary greatly and range between 10 and 500 ps. The

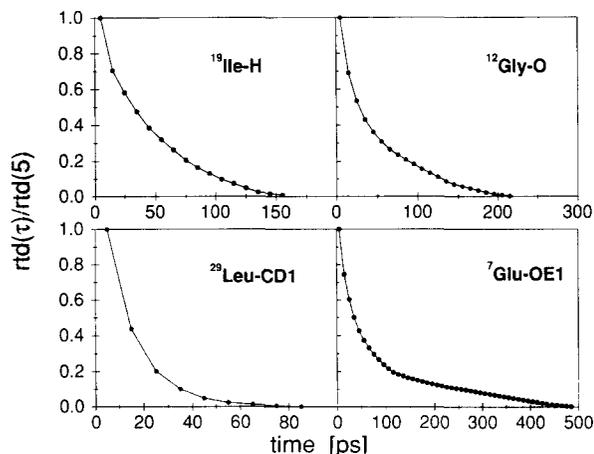


Fig. 16. Residence time distribution for water molecules in the first hydration shell of four atoms on the surface of BPTI as obtained from a 1 ns MD simulation of BPTI in aqueous solution [14].

average residence times are of the order of tens of picoseconds, which shows that the hydration shell of a protein is rather mobile.

### 2.3.7. Ion dynamics

When simulating a protein in aqueous solution, one may include counterions, such as  $\text{Na}^+$  or  $\text{Cl}^-$ , in order to obtain a neutral system or a specific ionic strength [26]. However, since hydrated ions diffuse rather slowly, the inclusion of ions in an MD simulation will lengthen the required equilibration and sam-

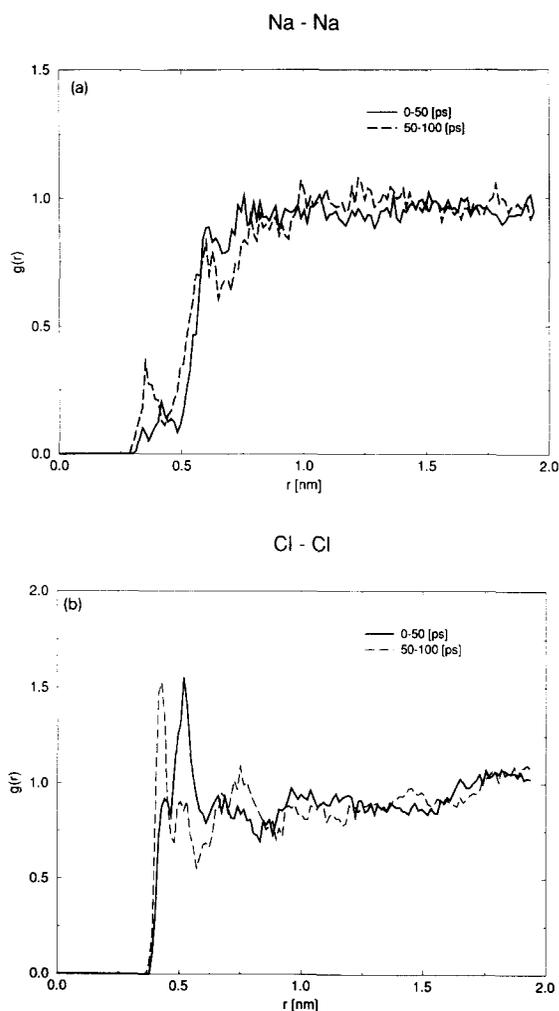


Fig. 17.  $\text{Na}^+ - \text{Na}^+$  (A) and  $\text{Cl}^- - \text{Cl}^-$  (B) radial distribution functions obtained from an MD simulation of a 1 molar sodium chloride solution [18] averaged over different 50 ps periods of the simulation.

pling time to far in the nanosecond time scale. The relatively long relaxation time of a distribution of ions in water is illustrated in Fig. 17, which shows the  $\text{Na}^+ - \text{Na}^+$  (A) and  $\text{Cl}^- - \text{Cl}^-$  (B) radial distribution functions obtained from an MD simulation of 40  $\text{Na}^+$  and 40  $\text{Cl}^-$  ions with 2127 water molecules in a cubic periodic box using the Ewald summation method to calculate the long-range electrostatic interactions [18]. The 50 ps averages differ considerably for different parts of the MD trajectory, which indicates that the ionic distribution has a relaxation time much longer than 50 ps.

### 3. Discussion and conclusions

The dynamics of proteins in solution is characterized by a variety of time scales, from femtoseconds to seconds. Of these time scales only the faster ones, up to nanoseconds, can currently be simulated on a computer at atomic detail because of the sizeable computing power that is required to integrate the atomic equations of motion over time. Local small-amplitude motions such as bond and bond-angle vibrations can be well sampled within tens of picoseconds. The dynamical properties of typical solvents such as water or chloroform can be simulated adequately with MD simulations covering hundreds of picoseconds. The residence times of water molecules at the protein surface lie in the range of tens to hundreds of picoseconds, which means that they can be crudely sampled in nanosecond simulations. Internal hydrogen bond fluctuations in a protein cover a wide range of time scales, of which only a part is within reach of simulations. It has been shown that a reliable reproduction of atomic B-factors derived from crystallographic refinement requires averaging periods of at least half a nanosecond for lysozyme. The dynamics of dielectric properties of proteins with sizeable contributions on a nanosecond time scale can only qualitatively be studied using nanosecond MD simulations.

A number of processes is still outside reach of even nanosecond MD simulations of proteins in water. This has been illustrated for the possible hinge-bending mode in lysozyme, for the translational and rotational diffusion of proteins and for the NMR  $^{15}\text{N}$  and  $^{13}\text{C}$  relaxation times. The dynamics of hydrated

ions in solution occurs on a nanosecond time scale, which makes a proper sampling of the counterion distribution in the solvent around a protein a difficult task.

The inaccessibility to simulation of the complete range of time scales contributing to a particular measurable protein property makes a proper comparison of simulated with measured values for such a property difficult. The observed discrepancy may be due to inadequate sampling of the relevant time scales, or to approximations in the atomic interaction function or invalid assumptions in the computational set-up of the simulation. Inadequate sampling of particular modes of motion can be detected by monitoring the time evolution of a variety of physical quantities, their mean values or fluctuations or by calculation of their correlation functions, if the length of the MD simulation is of the same size as the time scale of the particular mode. This has been illustrated for the calculation of atomic B-factors, NMR order parameters and free energies of complex formation. However, if the trajectory average of a particular quantity shows a stable average and fluctuations, one should be aware of the possibility that the observed stability is an artifact of the simulation, i.e. that it is solely due to a restriction of the motion of the protein imposed by the computational set-up. For example, the large surface tension of a protein in vacuo tends to frustrate a realistic simulation of the dynamics of a protein in solution by MD simulation in vacuo. The use of extended wall spherical shell boundary conditions will also severely restrict the amplitude of the atomic motion inside the spherical shell of restrained atoms. The use of instantaneous atom–atom distance restraints in protein structure refinement based on NMR data or the use of instantaneous structure factor restraints in such a refinement based on X-ray crystallographic data both have been shown to lead to a too rigid, static picture of a protein in solution or in crystalline environment [4,27,28].

The utility of MD simulation to study the dynamics of proteins lies mainly in the possibility to analyze at atomic detail the factors that are essentially contributing to a particular process or dynamical observable. This has been illustrated by the analysis of the relatively small contribution of the peptide dipoles to the dielectric relaxation of a protein, and

by the indication that for a simulation on a nanosecond time scale the  $T_1$  relaxation rates for  $^{15}\text{N}$  and  $^{13}\text{C}$  NMR relaxation are mainly determined by the size of the overall rotational correlation time and the order parameters of a protein in solution. Molecular dynamics simulation can definitely deepen our insight in the dynamics of proteins, be it that the simulated trajectories are interpreted with caution and a clear eye to the limitations of sampling on, currently, the nanosecond time scale.

## References

- [1] C.L. Brooks III, M. Karplus and B.M. Pettitt, *Proteins: a theoretical perspective of dynamics, structure, and thermodynamics*, Adv. Chem. Phys. Vol. 71 (Wiley, New York, 1988).
- [2] C.M. Dobson, P.A. Evans and S.E. Radford, *Trends Biochem. Sci.* 19 (1994) 31.
- [3] W.F. van Gunsteren in: *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*, Vol. 2, W.F. van Gunsteren, P.K. Weiner and A.J. Wilkinson, eds. (ESCOM, Leiden, 1993) p. 3.
- [4] W.F. van Gunsteren, R.M. Brunne, P. Gros, R.C. van Schaik, C.A. Schiffer and A.E. Torda, in: *Methods in Enzymology: Nuclear Magnetic Resonance*, Vol. 239, T.L. James and N.J. Oppenheimer, eds. (Academic Press, New York, 1994) p. 619.
- [5] M. Field, in: *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*, Vol. 2, W.F. van Gunsteren, P.K. Weiner and A.J. Wilkinson, eds. (ESCOM, Leiden, 1993) p. 82.
- [6] W.F. van Gunsteren and M. Karplus, *Macromolecules* 15 (1982) 1528.
- [7] H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, A. DiNola and J.R. Haak, *J. Chem. Phys.* 81 (1984) 3684.
- [8] T.C. Beutler and W.F. van Gunsteren, *Molecular Simulation* 14 (1994) 21.
- [9] H. Kovacs, A.E. Mark, J. Johansson and W.F. van Gunsteren, *J. Mol. Biol.* 247 (1995) 808.
- [10] R.M. Brunne and W.F. van Gunsteren, *FEBS Lett.* 323 (1993) 215.
- [11] J.E.H. Koehler, W. Saenger and W.F. van Gunsteren, *Eur. Biophys. J.* 16 (1988) 153.
- [12] P.E. Smith, R.M. Brunne, A.E. Mark and W.F. van Gunsteren, *J. Phys. Chem.* 97 (1993) 2009.
- [13] P.E. Smith and W.F. van Gunsteren, *J. Mol. Biol.* 236 (1994) 629.
- [14] R.M. Brunne, E. Liepinsh, G. Otting, K. Wüthrich and W.F. van Gunsteren, *J. Mol. Biol.* 231 (1993) 1040.
- [15] P.E. Smith, R.C. van Schaik, T. Szyperski, K. Wüthrich and W.F. van Gunsteren, *J. Mol. Biol.* 246 (1995) 356.
- [16] I.G. Tironi and W.F. van Gunsteren, *Mol. Phys.* 83 (1994) 381.
- [17] A.E. Mark, S.P. van Helden, P.E. Smith, L.H.M. Janssen and W.F. van Gunsteren, *J. Am. Chem. Soc.* 116 (1994) 6293.
- [18] I.G. Tironi, R. Sperb, P.E. Smith and W.F. van Gunsteren, *J. Chem. Phys.* 102 (1995) 5451.
- [19] W.F. van Gunsteren and H.J.C. Berendsen, *Groningen Molecular Simulation (GROMOS) Library Manual* (Biomos, Groningen, 1987).
- [20] M. Ramanadham, L.C. Sieker and L.H. Jensen, *Acta Crystallogr. A (Suppl.)* 43 (1987) 13.
- [21] B. Klar, B. Hingerty and W. Saenger, *Acta Crystallogr. B* 36 (1980) 1154.
- [22] W. Kabsch and C. Sander, *Biopolymers* 22 (1983) 2577.
- [23] A. Abragam, *The Principles of Nuclear Magnetism* (Clarendon, Oxford, 1961).
- [24] G. Lipari and A. Szabo, *J. Am. Chem. Soc.* 104 (1982) 4546.
- [25] H. Fröhlich, *Theory of Dielectrics* (Clarendon, Oxford, 1958).
- [26] J. de Vlieg, H.J.C. Berendsen and W.F. van Gunsteren, *Proteins* 6 (1989) 104.
- [27] A.E. Torda, R.M. Scheek and W.F. van Gunsteren, *J. Mol. Biol.* 214 (1990) 223.
- [28] C.A. Schiffer, P. Gros and W.F. van Gunsteren, *Acta Crystallogr. D* 51 (1995) 85.